Better Feared than Loved: Reputations and the Motives for Conflict

Iain W. Long
Cardiff University

April 2015

Abstract

Throughout history, victory in conflict has created fearsome reputations. Victory thus ensures greater allegiance of the wider population, increasing the victor's rents at the expense of their enemy. Such reputational concerns create two motives for conflict. When only the outcome is informative, the less feared party may attack to show that they are tougher than expected. If the fact that conflict occurred at all also conveys information, the more feared party may attack. If they do not, the population view peace as a sign of weakness and switch loyalties anyway. In this case, conflict arises to save face.
Keywords: Conflict, Reputations, Information Structures.
JEL: D74, C73, D83, F51, H56.

1 Introduction

> "My view is that it is desirable to be both loved and feared; but it is difficult to achieve both and, if one of them has to be lacking, it is much safer to be feared than loved."
>
> Machiavelli 1532, pp. 59.

Writing of the principalities of sixteenth century Italy, Machiavelli recommends that a prince should cultivate a fearsome reputation. Without it, his subjects' allegiance, and the wealth that they generate, may fall to somebody scarier. History is littered with examples of conflicts that arise because leaders wish to develop such notoriety at

the expense of a rival. These conflicts are not necessarily characterised by one group decisively crushing another, but rather an eventual settlement, determined in part by a third party.

The Roman Civil War of 49-45 B.C.E. arguably fits this pattern. Julius Caesar and Pompey the Great fought many battles, but it was ultimately down to a combination of the Roman Senate and the legion commanders which would be succeed. Victories increased the likelihood that the Senators would lend political support one party over the other. Similarly, legion commanders increasingly defected to Caesar as the war progressed.

Feudalism provides many rich examples. In addition to the situation outlined by Machiavelli, lesser nobles deferred to their liege lord to determine the territories that they controlled. The only way to improve their own position was to convince their lord that they were worthy of advancement. Worth was often determined by perceived military prowess, determined both on the battlefield and in competition (e.g. jousting).

Such reputational concerns are still very relevant to understanding contemporary conflict. The success of a coup d'état, or whether a counter-coup is instigated, often depends upon how fearful the population are of the new regime relative to the old (Jia and Liang 2014). Empirical evidence even suggests that insurgents have sabotaged aid programs, for fear that such programs will cause their support to wane (Crost et al. 2014). This not only demonstrates violent ability, but also highlights the importance of the wider population's support in a successful insurgency.

Criminologists have long agreed that creating a credible threat of violence is essential for modern criminal organisations. The territory or market share that a gang can hold onto is inextricably tied to the general population's perception of their toughness. Whilst this is particularly true of extortionists and protection rackets (Gambetta 1996; Konrad and Skaperdas 1997, 1998), it is also important for organisations selling illegal goods, such as drugs (Jankowski 1991; Levitt and Venkatesh 2000). In either setting, the perceptions of the gangs' customers are key. Which group they turn to for protection, or where they feel safest purchasing drugs will determine the revenues each gang receives. When deciding whether to attack a rival gang, each must weigh up the likely impact upon its reputation *ex post*.

I develop a new framework in which to evaluate such scenarios. Two parties have claims over a surplus generated by a population of atomistic observers in each of two periods. Whilst each party can be tough or weak, the share it receives does not depend upon ability. Instead, the observers lend their support (and hence their surplus) based upon how intimidating they perceive a party to be relative to its rival. Developing a

reputation for toughness, and ruining that of its rival, is thus each party's objective. Conflict enables them to do this. Whilst reputational concerns are clearly not the only motives for attacking a rival (see Garfinkel and Skaperdas 2007 for a review of the conflict literature), my aim is to identify conditions under which they do play a role.

In each period, both parties choose whether to attack their rival. If nobody attacks, observers learn that peace prevails. They may then revise both parties' reputations, and the surplus is shared out accordingly. If, on the other hand, one or both parties attack then conflict breaks out. Conflict has two features. Firstly, the current surplus is destroyed in the fighting. Secondly, one party is declared the victor, while the other is defeated. Tough parties are more likely to be victorious, and so the victor's reputation improves whilst their rival's declines. Moreover, the mere fact that fighting broke out may reveal information in equilibrium. This causes further reputational adjustments.

Whilst observers learn that conflict broke out, they cannot tell who initially attacked (although they will learn this almost surely in equilibrium). Deception has a long history in conflict, dating back to at least 2500 years (Sun Tzu c. 500 B.C.E., chapter 1, verses 18-20). It is often impossible for observers to know who fired the first shot. In more modern times, the use of propaganda in the run up to conflict is well documented (c.f. Glaeser 2009; Edmond 2013). Each side tries to convince the population that they have been unjustly provoked by their rival. In this setting, such messages would constitute cheap talk, although they are not explicitly modelled.

When deciding whether to attack, both parties act as though their actions will prove pivotal. If a party is attacked, its own action does not affect the outcome of the ensuing conflict. Each consequently focuses on the case where what they do matters, i.e. when its rival does not attack. In the final period, this implies that peace prevails. Neither party has further reputational concerns, and does not wish to lose its share of the current surplus.

Knowing that peace will prevail in the future, both parties have reputational concerns in period one. For conflict to be worthwhile, the attacker must expect to receive a small share of the first period surplus should peace prevail. It has little to lose by fighting. Conversely, its rival must receive the lion's share, making conflict prohibitively costly for them. Only the party with the worse peacetime reputation ever attacks. Upon observing conflict, the observers can immediately infer who attacked. Deception does not occur in practice.

I compare two information structures. In the first, abilities are unknown to the parties. Although each is likely to know its own military ability, there may be uncertainty regarding the distribution of abilities. Parties are therefore unsure whether the observers

would view them as tough or weak if their ability were revealed. Each shares the prior beliefs of the observers. Since actions carry no information, either peace prevails in both periods, or the party with the weaker reputation attacks in period one. In this case, conflict is caused by the weaker party expecting to improve its reputation, and hence its future payoff, by attacking. Conflict occurs to prove people wrong.

I then allow ability to be private information. Whilst it is still possible that attacks occur to prove people wrong, new equilibria become possible. Since a tough party is more likely to be victorious, it always enjoys a higher expected payoff from attacking than a weak party. If the breakout of conflict conveys information, observers must upgrade the reputation of the likely attacker when a fight occurs. If peace prevails, on the other hand, it is more likely that the would-be attacker is weak. This provides an alternative motive for conflict. Rather than expecting a reputational gain from fighting, a party may attack to prevent a reputational loss from maintaining the peace. Conflict occurs to prevent the loss of face.

The remainder of the paper continues as follows. In the next section, I place the current work in the broader literatures on conflict and on reputations. Whilst my contributions are primarily to the conflict literature, I make extensive use of the tools developed to analyse reputations. In Section 3, I outline my model. Section 4 begins by identifying some general results which hold for any information structure. Sections 4.3 and 4.4 then discuss equilibria in the case where parties do not know their own ability and when it is private information respectively. Finally, Section 5 concludes. Proofs are presented in the appendices.

2 Related Literature

Conflict over resources has long been of interest to economists (since, at least, Tullock 1967). Unlike market allocations, those deriving from conflict generate inefficiencies. Resources may be destroyed or diverted from productive to acquisitive activities. Parties with larger military arsenals are more likely to be victorious, providing an incentive to invest in guns (Tullock 1980; Garfinkel 1990; Grossman 1991; Hirshleifer 1991; Grossman and Kim 1995). The marginal cost of this investment is a reduction in the quantity of consumer goods produced, making victory less profitable. Whilst this approach has strong foundations (c.f. Hirshleifer 1989; Skaperdas 1996), I take a more stylised route. In the forthcoming analysis, it is only important that victory signals intrinsic toughness to the observers. As such, I treat parties' stocks of weaponry and military know-how as given. They are either tough or weak. In this sense, the model represents a short-run

analysis. Resource destruction, which still acts as a counterbalance to the incentive to attack, is assumed to be complete in any period in which conflict occurs. Whilst not necessary, it does simplify the analysis to allow clearer discussion of the underlying economics.

Mine is not the first paper to identify the importance of information in generating conflict (see Fearon 1995 for a literature review). Whilst some have considered uncertainty regarding the common value of the resources being fought over (Wärneryd 2003; Chassang and Padró i Miquel 2010), other have treated military ability as unknown or private information (Brito and Intriligator 1985; Bueno de Mesquita et al. 1997). I take the latter approach. Conflict can arise because both parties are too optimistic regarding how likely it is that they will be victorious. As such, it may be the case that they cannot reach an agreement through negotiation. These situations also create an incentive to signal. Two recent papers have considered signalling leading to arms races (Baliga and Sjöström 2004) and prevention of weapons inspection (Baliga and Sjöström 2008). Silverman 2004 goes one step further, analysing a random-matching environment in which individual violent behaviour signals toughness. By gaining a reputation for violence, an individual can ensure that rivals behave passively towards them. I add to this literature, as both actions and outcomes have the potential to signal ability. Moreover, I allow payoffs to depend directly upon reputations, by introducing outside observers.

The destruction of resources would suggest that ex ante negotiation would be more efficient. However, such negotiation must be credible. Recently, conflict in a dynamic setting has been investigated. When parties interact repeatedly, threats of future conflict can lead to the signing of a social contract today (Garfinkel 1990; Azam and Mesnard 2003; Bester and Wärneryd 2006). Alternatively, longer time horizons can generate more conflict. Victory today can lead to higher payoffs in the future (Garfinkel and Skaperdas 2000). As such, costly conflict can be thought of as an investment in future gains. My contribution is much closer in nature to this latter work. Parties have reputational concerns. They may, therefore, be willing to forego their share of the surplus today for the possibility of a greater share of future surpluses. Whilst ex ante negotiation is not possible due to the short time horizon I consider, one could reinterpret my results as identifying where the threat of conflict is credible, should negotiations break down.

To the best of my knowledge, this is the first analysis to consider the signals generated by conflict (or, for that matter, by peace). This creates different incentives for the parties involved; those associated with reputations. Unlike previous models in the literature, it is observers' beliefs about the intrinsic toughness of each party, rather than toughness itself, that determines their payoffs.

The direct dependence of payoffs upon the beliefs does, however, create problems. In deciding what to do, parties must know exactly how each possible action will be interpreted. But if the action under consideration is completely unexpected (i.e. never occurs in equilibrium), any observer belief is consistent. These models have the potential to be plagued by innumerable equilibria, based on 'crazy' out of equilibrium beliefs. In order to focus on more meaningful equilibria, the reputations literature often introduces myopic players (Kreps and Wilson 1982; Milgrom and Roberts 1982). With a small probability, each player always takes the same action in every period. Since each action is allotted a myopic type, every action is observed with positive probability in every period. Since Bayes's rule can always be applied, there is a unique set of beliefs consistent with any action. I adopt a similar approach, by introducing agency problems. With some small probability the orders given by each party are either misinterpreted (as it is believed they were at the charge of the Light Brigade) or ignored. Hence, even if neither party is expected to attack, there is a unique way to interpret conflict.

These models commonly feature one long-run player facing a series of short-run players (see also Fudenberg and Levine 1989, 1992).[1] Importantly, it is only the long-run player who has reputational concerns. Each short-run player is unable to make use of the information they acquire during their interactions with the long-run player. My model is different. Reputational externalities exist between the two parties. By choosing to attack, a party not only affects their own reputation, but also that of their rival. Having two parties with private information also creates the possibility of disagreement regarding each parties' reputation. By truthfully announcing its beliefs about the ability of its rival, a party could reveal its own ability. This creates an incentive for a weak party to mimic a tough party. I get around such problems by introducing outside observers, whose beliefs do not depend upon any private information, and hence can be agreed upon by all groups.

Reputations have been successfully applied to models of product quality. Consumers' willingness to pay is determined by the expected quality of the product that they are purchasing - reputation has value (Shapiro 1982; Tadelis 1999, 2002; Bar-Isaac 2003). My model shares this feature, insofar as reputations enter directly into parties' payoffs. I also borrow structure from the latter papers. I first assess behaviour under the assumption that parties do not know their own ability, before making ability private information.

[1]Even when considering two long-run players, the type of one player may be assumed to be common knowledge. Only one side has reputational concerns (Atakan and Ekmekci 2012).

3 A Model of Conflict and Reputation

Two parties, $k \in \{i, j\}$, share a surplus, $S_t \in \{0, 1\}$, in each of two periods. Both parties are characterised by an innate ability, θ^k. They can either be *tough* ($\theta^k = 1$) or *weak* ($\theta^k = 0$). Whilst ability is important in determining the outcome of conflict (more on this below), it does not directly impact upon the share of the surplus each party receives. Instead, this is determined by a mass of atomistic observers. Observers do not know each party's ability, and instead form a belief - the parties' reputations, $\mu_t^k = \mathbb{E}_t\left[\theta^k\right] = \mathbb{P}_t\left(\theta^k = 1\right)$. It is these reputations that determine the share of the surplus that each party receives. Observers' priors regarding party k's ability are μ_0^k. As information is revealed during the game, observers will update reputations accordingly.

In each period, parties either attack their rival ($a_t^k = 1$) or not ($a_t^k = 0$). If at least one party attacks, conflict occurs. Conflict has two features. Firstly, it destroys the current period's surplus: $S_t = 0 \iff \max\{a_t^i, a_t^j\} = 1$. Secondly, it has a commonly known *outcome*: a winner and a loser. The probability that a particular party is victorious depends upon the how its ability compares to that of its rival. If both share the same ability, each has an equal probability of winning. If a tough party fights a weak one, the tough party is victorious with probability $\frac{1}{2} + q$ for some $q \in \left(0, \frac{1}{2}\right]$.

Whilst conflict destroys today's surplus, it also sends a signal to the observer regarding each party's ability. I assume that the observers cannot tell who starts the conflict[2], however they do observe the outcome. This allows for deception or the use of propaganda. It also simplifies the analysis significantly. In response, they update the parties' reputations in two stages. Taking each parties' *prior reputation*, μ_{t-1}^k, as given, whether conflict occurs may tell the observers something. Based on this, they develop an *interim reputation* by applying Bayes's rule. Secondly, conditional on conflict occurring, the outcome reveals new information. Since victory is positively correlated with being tough, the victor's reputation improves. Similarly, the reputation of the loser declines. Again applying Bayes's rule, the observers develop *posterior reputations*, μ_t^k, for each party.

I distinguish between two *information structures*. Under the first, *unknown abilities*, the parties share the observers' beliefs regarding their own ability. This may reflect uncertainty not about their own military capacity, but rather how it compares to that of their rival. If a comparison is difficult, the party simply believes its own reputation. Under the second information structure, *private abilities*, comparison is possible. Each party knows its own military capacity, as well as the distribution of its rival's capacity. It

[2]In any equilibrium, they will learn this almost surely anyway.

is thus in a position to know its own ability. I assume that each still share the observers' priors about the ability of its rival at the start of the game.

The proportion of the surplus party i receives in period t is given by a *sharing rule*, $\pi(\mu_t^i, \mu_t^j) \in [0,1]$. As I only consider the decisions of the parties, the sharing rule is treated as both exogenous and common knowledge. I assume that π satisfies:

1. **Symmetry:** $\pi(\cdot,\cdot)$ is common to both parties.
2. **Efficiency:** All the surplus is shared. $\pi(\mu_t^i, \mu_t^j) + \pi(\mu_t^j, \mu_t^i) = 1$.
3. **Gains to reputation:** A party's share is increasing in his own reputation, unless they are already receiving the entire surplus. $\pi(\cdot,\cdot)$ is increasing in its first argument.

Symmetry ensures that a party's payoff is independent of his identity. Only the reputations matter. Efficiency ensures that there is no waste in the sharing of the surplus. The game is strictly competitive. Gains from reputation states that an increase in reputation raises the share that the party expects to receive. Combined with efficiency, the share a party expects is clearly declining in the reputation of its rival. Readers familiar with the contests literature will realise that the three conditions are identical to those used to axiomatise a contest success function (Skaperdas 1996). However, the interpretation is different. Contests involve employing resources to increase the probability of success (Tullock 1980). This sharing rule requires no additional resources to be expended. Instead, it depends upon the population's perceptions about both parties. Whilst it still states the share of the surplus each party can expect, it does not entail the waste of resources imagined by Tullock (1967). Party i's discounted utility is:

$$\pi(\mu_1^i, \mu_1^j)S_1 + \delta\pi(\mu_2^i, \mu_2^j)S_2, \tag{1}$$

where the discount rate is $\delta \in (0,1)$.

In order to illustrate results later in the paper, it is useful to have an example sharing rule:

Example (Nash Bargaining) *The share party i receives of the surplus in period t is given by the solution to:*

$$\max_{\pi\in[0,1]} \left\{ (\pi S_t)^{\mu_t^i} [(1-\pi)S_t]^{\mu_t^j} \right\},$$

which yields:

$$\pi(\mu_t^i, \mu_t^j) = \frac{\mu_t^i}{\mu_t^i + \mu_t^j}.$$

Clearly, this satisfies the three conditions defining a sharing rule. The Nash bargaining solution is always efficient. Party j receives a share $1-\pi(\mu_t^i,\mu_t^j)=\pi(\mu_t^j,\mu_t^i)=\frac{\mu_t^j}{\mu_t^i+\mu_t^j}$*, so the sharing rule is also symmetric. Finally, the share each party receives is increasing in its own reputation, unless its rival is revealed to be weak. In the latter case, the party receives the entire surplus, and so gains from reputation is also satisfied. Of course, the Nash bargaining solution is equivalent to the power contest success function (Hirshleifer 1989).*

The dependency of payoffs upon the observers' beliefs create a common problem in the reputations literature. Suppose parties are in an equilibrium in which neither one attacks. Observers have no way to interpret conflict, as it conditions on a zero probability event. Since these beliefs affect parties' incentive to deviate, it is possible to support a range of equilibria simple by constructing 'crazy' out-of-equilibrium beliefs. In order to preclude these possibilities, I introduce an agency problem. Each party can be thought of as an army. A general gives the orders, and the soldiers follow them. With some small probability, $\varepsilon^k \in (0,1)$, the general's orders are either lost or ignored. In this case, the soldiers attack with probability $\eta^k \in (0,1)$. Even in the equilibrium in which neither party chooses to attack, conflict occurs with positive probability. There is a unique way for the observers to update reputations.

For notational ease, I will consider the limiting case when $\varepsilon^k \to 0$ (in a similar manner to Tadelis 1999, 2002). More formally, this represents a refinement of extensive form trembling hand perfection (Selten 1978). Extensive form trembling hand perfection requires the existence of some sequence of trembles for which the equilibria of the perturbed game converge to the equilibrium in the unperturbed game. By changing the relative sizes of these trembles for tough and weak parties, it is possible to construct any arbitrary out-of-equilibrium beliefs, including 'crazy' ones. To rule them out, I focus on situations where both abilities of a given party tremble with the same probability. Intuitively, this ensures that observers do not read too much into mistakes, and do not drastically change their beliefs in light of something unexpected.

The timing in the model is as follows. At the start of each period, prior reputations are common knowledge. Depending upon the information structure, parties may also know their own ability. Both parties simultaneously choose whether to attack or not. They may then suffer agency problems. If at least one attacks, conflict breaks out. The current-period surplus is destroyed. The observers note the conflict, and generate interim reputations for each party. The outcome of the conflict is then revealed. Observers generate posterior reputations. If neither party attacks (either through choice or as a result of agency problems), then the observer notes the peace and generates posterior

reputations for both parties. The surplus is shared out accordingly. If there is another period to play, posterior reputations in the current period become prior reputations in the following period.

4 Equilibria

An equilibrium in this model is defined as follows:

Definition 1 *Given the observers' priors $\mu_0 = (\mu_0^i, \mu_0^j)$, a set of strategies (attack probabilities) for both parties, $\sigma_t = (\sigma_t^i, \sigma_t^j)$ for $t = 1, 2$, and a set of posterior reputations, $\mu_t = (\mu_t^i, \mu_t^j)$ for $t = 1, 2$, constitute an equilibrium if and only if:*

1. *In each period both parties' strategies maximise their expected payoff:*

$$\sigma_1^k(\theta^k) \in \arg\max_{\sigma\in[0,1]} \left\{ \mathbb{E}_1 \left[\pi(\mu_1^k, \mu_1^{-k}) S_1 + \delta \pi(\mu_2^k, \mu_2^{-k}) S_2 | q, \mu_0, \sigma_t^{-k}, \mu_t, t = 1, 2 \right] \right\},$$
$$\sigma_2^k(\theta^k) \in \arg\max_{\sigma\in[0,1]} \left\{ \mathbb{E}_2 \left[\pi(\mu_2^k, \mu_2^{-k}) S_2 | q, \mu_1, \sigma_2^{-k}, \mu_2 \right] \right\},$$

   for each $k \in \{i, j\}$ and each $\theta^k \in \{0, 1\}$ (if party k knows θ^k).

2. *Both parties and the observers generate posterior reputations by applying Bayes's rule, given the strategies of both parties.*

Each party maximises their payoff in both periods, given their information, the strategy of their rival and the way that the observers update the reputations. As information is revealed, both the parties and the observers update their beliefs accordingly. Since each party faces an agency problem, there is always a unique way to do this, applying Bayes's rule.

In characterising the equilibria of this game, I will first show how reputations are updated in more detail, before turning to the strategies of the parties.

4.1 Updating Reputations

Let $\sigma_t^k(\theta) \in \left[\eta^k \varepsilon^k, 1 - (1 - \eta^k)\varepsilon^k\right]$ denote the equilibrium probability that party k with ability θ attacks. Suppose first that peace prevails. The observers learn that neither

party attacked. They generate an interim reputation for party k according to:

$$\begin{aligned}\mu_t^{kP} &= \mathbb{P}\left(\theta^k = 1 | \text{Peace in } t\right) \\ &= \frac{\mu_{t-1}^k \left[1 - \sigma_t^k(1)\right]}{\mu_{t-1}^k \left[1 - \sigma_t^k(1)\right] + (1 - \mu_{t-1}^k)\left[1 - \sigma_t^k(0)\right]}.\end{aligned} \tag{2}$$

If conflict did not occur, it must be the case that party k did not attack. Conditional on being tough, it does not attack with probability $1 - \sigma_t^k(1)$. The numerator of (2) is $\mathbb{P}\left(\theta^k = 1\right)\mathbb{P}\left(a_t^k = 0 | \theta^k = 1\right)$, the probability of being tough and not attacking. The denominator is the total probability that party k does not attack. Note that, if a tough party k is more likely to attack than a weak party k, i.e. that $\sigma_t^k(1) > \sigma_t^k(0)$, then peace reduces the party's reputation.

If peace prevails, party k's posterior reputation is identical to its interim one: $\mu_t^k = \mu_t^{kP}$. No further information is revealed. Parties i and j receive shares $\pi(\mu_t^{iP}, \mu_t^{jP})$ and $\pi(\mu_t^{jP}, \mu_t^{iP})$ respectively. They not only enjoy these shares in the future but, by preserving the peace, they also receive a payoff in the current period. They share a total discounted surplus of $1 + \delta$.

Now suppose that conflict occurs. The observers know that at least one party attacked, although due to the possibility of deception or the use of propaganda, they are unsure who initiated hostilities. It could be, for example, that party i attacked (in which case the action of j would be irrelevant) or that it did not attack but was attacked by j. The observers consequently develop an interim reputation for party i as follows:

$$\begin{aligned}\mu_t^{iC} &= \mathbb{P}\left(\theta^i = 1 | \text{Conflict in } t\right) \\ &= \frac{\mu_{t-1}^i \left\{\sigma_t^i(1) + \left[1 - \sigma_t^i(1)\right]\sigma_t^j\right\}}{\mu_{t-1}^i \left\{\sigma_t^i(1) + \left[1 - \sigma_t^i(1)\right]\sigma_t^j\right\} + (1 - \mu_{t-1}^i)\left\{\sigma_t^i(0) + \left[1 - \sigma_t^i(0)\right]\sigma_t^j\right\}},\end{aligned} \tag{3}$$

where $\sigma_t^j = \mu_{t-1}^j \sigma_t^j(1) + (1 - \mu_{t-1}^j)\sigma_t^j(0)$ is the probability that party j attacks party i. The numerator of (3) gives the probability that i is tough and conflict occurs. With probability μ_{t-1}^i, i is tough. In this case, conflict will occur when i attacks (which happens with probability $\sigma_t^i(1)$), irrespective of the action of j. If i does not attack (which happens with probability $1 - \sigma_t^i(1)$), then conflict only occurs if j attacks. The denominator gives the total probability of conflict. Party j's interim reputation is generated in an identical manner. Note that, if a tough party k is more likely to attack than a weak party k then the fact that conflict occurs improves the party's reputation.

Should conflict occur, the outcome reveals more information. Tough parties are more likely to be victorious, and so the outcome signals ability. Given the interim reputations,

however, this has no strategic element. It is purely a lottery. From the perspective of the observers, for a given interim reputation for party j, a tough party i is victorious with probability:

$$\begin{aligned}\mathbb{P}\left(i \text{ is victorious in } t|\theta^i=1\right) &= \tfrac{1}{2}\mu_t^{jC} + \left(\tfrac{1}{2}+q\right)\left(1-\mu_t^{jC}\right) \\ &= \tfrac{1}{2} + q\left(1-\mu_t^{jC}\right), \end{aligned} \tag{4}$$

They win half of the time if party j is tough, and with probability $\frac{1}{2}+q$ if j is weak. A weak i, on the other hand, is only victorious with probability:

$$\begin{aligned}\mathbb{P}\left(i \text{ is victorious in } t|\theta^i=0\right) &= \left(\tfrac{1}{2}-q\right)\mu_t^{jC} + \tfrac{1}{2}\left(1-\mu_t^{jC}\right) \\ &= \tfrac{1}{2} - q\mu_t^{jC}. \end{aligned} \tag{5}$$

The higher probability of victory of a tough i causes the observers to improve their estimation of i's ability. After the announcement of the outcome of the conflict, the observers develop a posterior reputation if i was victorious as follows:

$$\begin{aligned}\mu_t^{iV} &= \mathbb{P}\left(\theta^i=1|i \text{ was victorious in } t\right) \\ &= \frac{\mu_t^{iC}\left[\frac{1}{2}+q\left(1-\mu_t^{jC}\right)\right]}{\mu_t^{iC}\left[\frac{1}{2}+q\left(1-\mu_t^{jC}\right)\right] + \left(1-\mu_t^{iC}\right)\left(\frac{1}{2}-q\mu_t^{jC}\right)}. \end{aligned} \tag{6}$$

The numerator states the probability that i is tough and victorious. The denominator gives the overall probability of victory. Note that victory causes party i's reputation to improve: $\mu_t^{iV} > \mu_t^{iC}$.

If, on the other hand, party i is defeated:

$$\begin{aligned}\mu_t^{iD} &= \mathbb{P}\left(\theta^i=1|i \text{ was defeated in } t\right) \\ &= \frac{\mu_t^{iC}\left[\frac{1}{2}-q\left(1-\mu_t^{jC}\right)\right]}{\mu_t^{iC}\left[\frac{1}{2}-q\left(1-\mu_t^{jC}\right)\right] + \left(1-\mu_t^{iC}\right)\left(\frac{1}{2}+q\mu_t^{jC}\right)}. \end{aligned} \tag{7}$$

The interpretation is broadly similar to that of (6). Since i is more likely to lose when weak, $\mu_t^{iD} < \mu_t^{iC}$.

The value of conflict is now apparent. By fighting and winning, a party enjoys a double-dividend. Its own reputation improves. Since there are gains to reputation, this increases the share of the surplus it enjoys given the reputation of its rival. However, the

rival has also been defeated, causing its reputation to decline. Gains to reputation imply that the rival's share of the surplus falls. Efficiency means that the party enjoys a further boost to the surplus it receives. Shares are $\pi(\mu_t^{iV}, \mu_t^{jD})$ and $\pi(\mu_t^{jD}, \mu_t^{iV})$ respectively. In this sense, fighting is an investment in a party's future reputation. Of course, victory is not guaranteed. It is also possible that he loses. The investment is risky. The cost of the investment is the destroyed surplus in the current period.

4.2 Parties' Behaviour

I now turn to the behaviour of the parties, given the beliefs of the observers. To begin with, I present general results which hold for either information structure. Instead, they hinge on either agency problems or surplus destruction. The first of these states how parties assess the payoffs from attacking or not in any equilibrium:

Proposition 1 (Pivotal Attacks) *For any sharing rule and either information structure, parties choose whether to attack in each period by conditioning both on not suffering an agency problem and on their rival not attacking.*

Proof. See A. ■

In the event that its rival attacks, a party's own action is irrelevant. Conflict occurs, and the observers do not see who started it. Similarly, if a party suffers agency problems, the original action it chose has no bearing on whether conflict occurs. It is only when its rival does not attack (which always occurs with positive probability due to their own agency problems) and when the party does not suffer agency problems (which again occurs with positive probability) that their action becomes relevant. When choosing whether to attack, each party focuses on this case. They behave as though their actions were pivotal.

Knowing how parties assess costs and benefits, I can now start to describe their actions in equilibrium. Parties are forward-looking, and so I apply backwards induction. Starting with the final period, the unique actions of each party follow immediately from Proposition 1. In period two, neither one has any concerns about the future. If either party attacks, the surplus is destroyed. Neither receives anything. However, if peace prevails, the surplus is maintained, and both receive strictly positive payoff. Defending is a dominant action:

Corollary 1 (No Conflict in Period Two) *For any sharing rule and either information structure, conflict only occurs in the second period as the result of agency problems.*

Proof. See B. ■

Suppose that party i knows that party j was able to attack with probability one and was intending to do so. Irrespective of i's action, conflict occurs and the second period surplus is destroyed. i is indifferent between attacking and not, and might as well choose to attack. In the unperturbed game, it is possible to support an equilibrium in which both parties attack for certain, purely because they know that their rival will attack anyway. Agency problems rule this out. Even if a party's rival chooses to attack, there is always a small probability that its orders will be lost, and it will not attack. This is the only case in which the party's action becomes payoff-relevant, and so it optimally choose not to attack. By doing so, the surplus is preserved. As both parties behave identically, conflict can only occur due to agency problems.

We now turn to period one. Since parties expect to receive a positive surplus in period two, both have reputational concerns. By choosing to attack, they give up any surplus in the current period. However, they may improve their reputation, leading to them receiving a larger share of the surplus in the final period. However, it can never be the case that both parties find such a strategy profitable:

Proposition 2 (Only One Party Ever Attacks) *For any sharing rule and either information structure, if one party chooses to attack in period one with positive probability, then its rival will always choose not to attack, irrespective of the rival's ability.*

Proof. See C. ■

Conflict in the first period destroys the surplus. For it to be profitable to attack, it must be the case that the attacker receives a relatively small payoff if peace prevails. In particular, it must be less than δ, the maximum discounted payoff a party can receive from conflict. However, as the sharing rule is efficient, its rival must receive a relatively large share of the period one surplus. For them, conflict is prohibitively costly. If peace prevails, the parties share $1+\delta$. If one party receives less than δ, their rival must receive a payoff greater than one. Since it is always possible that the attacker suffers agency problems, its rival strictly prefers not to attack, hoping to enjoy a large share of the surplus in both periods. Of course, this result hinges on the assumption that $\delta < 1$. Whilst it greatly simplifies the analysis, the motives for conflict identified below would not change if this assumption were relaxed.

One implication of Proposition 2 is that the observer learns the identity of the attacker almost surely. For it to be worthwhile for one party to attack, the other must strictly prefer peace. Upon observing conflict, only the party who is expected to attack sees a change in their interim reputation. The other party is assumed to have opted

not to attack. In principle, this could allow for deception. By provoking conflict, the supposedly peaceful party could trick the observers into believing that its rival had attacked. However, it would never have an incentive to do so. The destruction of resources would always be too costly.

A second implication is that all the results that follow would hold if parties were to have different information, i.e. if a party with unknown ability were to face a party with private ability. Since only one party would ever attack in equilibrium, the knowledge of its rival is irrelevant. Not attacking is a dominant action.

The intuition underpinning Proposition 2 allows us to go slightly further, and identify the party who may have an incentive to attack:

Corollary 2 (Better Peacetime Reputation Never Attacks) *For any sharing rule and either information structure, the party with the (weakly) better reputation when peace prevails in period one will always choose not to attack.*

Proof. See D. ■

The party with the weaker peace time reputation has the least to lose from conflict. It receives a smaller share of the period one surplus. As such, it is the one who is likely to attack. The party with the tougher peace time reputation, on the other hand, receives a very high payoff from maintaining the peace - the majority of the surplus in both periods. It has the greater interest in maintaining the status quo, and hence never chooses to attack.

In what follows, we can therefore restrict attention to situations in which $\mu_1^{iP} < \mu_1^{jP}$, and j never attacks. We then need to explore i's decisions. The case when $\mu_1^{iP} > \mu_1^{jP}$ will be entirely symmetric, only i will never attack. Of course, if $\mu_1^{iP} = \mu_1^{jP}$, then the parties share the surplus evenly, each receiving $\frac{1}{2}(1+\delta)$. Since $\delta < 1$, neither party has an incentive to attack. Reputational gains can never compensate for the loss of period one surplus. We can never have conflict when two parties have the same peacetime reputation, except as the result of agency problems.

In the case where abilities are unknown, this is equivalent to focusing on $\mu_0^i < \mu_0^j$. However, as well shall see, the situation with private abilities is more subtle. Peace reveals information about both parties. After all, observers learn that neither attacked. This could lead to situations in which a party with a tougher prior reputation finds itself with the weaker posterior reputation when peace prevails. This intuition will enable us to characterise equilibria which appear at odds with the intuition behind Corollary 2; ones in which the party who seems tougher at the start of the period will choose to attack.

Note that, in this simple framework, credible commitment is not possible. Suppose that party i wished to pay party j not to attack in period one. Since the surplus only survives if peace prevails, i would only be able to pay j after actions had been chosen. But then, knowing that j would never attack in period two, i would have an incentive to withhold the promised funds. Of course, one could artificially create the possibility of a binding social contract. In this case, the following results could be reinterpreted as conditions under which the threat of conflict is itself credible.

4.3 Unknown Abilities

In order to make further progress, I now consider the equilibria under the assumption that parties are not informed about their own abilities. As discussed previously, this could represent the case in which ability is a relative measure. Whilst they know how capable they are in absolute terms, they are unclear regarding the exact definitions of tough or weak. In this case, they share the observers' beliefs regarding their abilities.

When abilities are unknown, parties cannot condition their actions upon them. Hence, whether a party attacks yields no information. Interim reputations are identical to prior ones (in both (2) and (3), $\sigma_t^k(0) = \sigma_t^k(1)$). Let us assume, without loss of generality that $\mu_0^i < \mu_0^j$. From Corollary 2, party j would never attack. We are therefore free to focus on the decision of party i.[3] If peace prevails, no information has been revealed about either party. Conditional on being pivotal, i expects to receive:

$$\pi(\mu_0^i, \mu_0^j) + \delta\pi(\mu_0^i, \mu_0^j).$$

If, instead, conflict occurs, the outcome still reveals something about the relative abilities of the parties. When it is announced, both the parties and the observer will use (6) and (7) to develop posterior reputations. Since party i does not know their own ability, it believes that it will be victorious with probability:

$$\begin{aligned}\mathbb{P}\left(i \text{ victorious}|\mu_0^i, \mu_0^j\right) &= \mu_0^i\left[\tfrac{1}{2} + q\left(1 - \mu_0^j\right)\right] + \left(1 - \mu_0^i\right)\left(\tfrac{1}{2} - q\mu_0^j\right) \\ &= \tfrac{1}{2} + q\left(\mu_0^i - \mu_0^j\right). \end{aligned} \tag{8}$$

If party i were to attack, it would therefore expect to receive:

$$0 + \delta\left\{\left[\tfrac{1}{2} + q\left(\mu_0^i - \mu_0^j\right)\right]\pi(\mu_1^{iV}, \mu_1^{jD}) + \left[\tfrac{1}{2} - q\left(\mu_0^i - \mu_0^j\right)\right]\pi(\mu_1^{iD}, \mu_1^{jV})\right\}.$$

[3] Of course, if $\mu_0^j < \mu_0^i$, the roles would simply be reversed.

In period one, the surplus is destroyed. In period two, neither party attacks, and so the surplus is preserved. If i is victorious, it will enter period two with a stronger reputation and will receive $\pi(\mu_1^{iV}, \mu_1^{jD})$. If, on the other hand, it is defeated, it will enter period two with a weaker reputation and will receive $\pi(\mu_1^{iD}, \mu_1^{jV})$.

Party i will therefore choose to attack if and only if:

$$\pi(\mu_0^i, \mu_0^j)(1+\delta) \leq \delta \left\{ \left[\tfrac{1}{2} + q\left(\mu_0^i - \mu_0^j\right)\right] \pi(\mu_1^{iV}, \mu_1^{jD}) + \left[\tfrac{1}{2} - q\left(\mu_0^i - \mu_0^j\right)\right] \pi(\mu_1^{iD}, \mu_1^{jV}) \right\}. \tag{9}$$

The incentive to attack is generated by the possibility of *proving people wrong*. Since attacking destroys the first period surplus, a party would only do so if it expected to gain a larger share of the surplus in the second period. Since its share under peace is determined by its prior reputation, this requires that they expect their reputation to improve after the conflict. We have the following:

Proposition 3 (Proving People Wrong) *When abilities are unknown, the unique equilibrium involves either peace in both periods, or the party with the weaker prior reputation attacking to prove people wrong in period one.*

The exact nature of the equilibrium depends upon the functional form of the sharing rule. Figure 1 illustrates it under Nash bargaining. The diagram shows the equilibrium outcome for any possible combination of prior reputations. A party will attack if and only if its reputation is low enough, compared to that of its rival. In this case, it receives too small a share of the first period surplus to prevent it from choosing to attack.

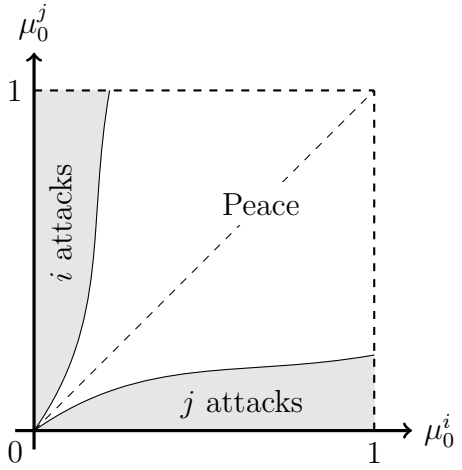

Figure 1: Peace and Proving People Wrong under Nash Bargaining with Unknown Abilities

Parties with identical reputations never attack one another (the dashed diagonal line). Each finds the loss of period one surplus too costly.

4.4 Private Abilities

Under this information structure, parties know their own ability, but share the observers' prior beliefs about their rival. The existence of private information enables the break out of conflict to signal information about the ability of the attacker. This information will always lead the observers to adjust reputations as follows:

Proposition 4 (Conflict Signals Toughness) *For any sharing rule, if a party is expected to attack with positive probability, attacking (weakly) improves their interim reputation, whereas not attacking (weakly) reduces it.*

Proof. See E. ■

The payoff a party receives should peace prevail in both periods is independent of ability. Similarly, should conflict break out, the way that interim (and hence posterior) reputations are updated is the same irrespective of whether the attacking party is tough or weak. Given any outcome of conflict, tough and weak parties thus receive the same shares of the remaining surplus. However, tough parties are more optimistic about their chances of being victorious than weak ones. They are consequently more prone to attacking. If a weak party chooses to attack, its expected payoff from attacking must exceed that of peace. Consequently a tough party's expected payoff from attacking strictly exceeds that of peace. Similarly, if a tough party chooses not to attack, its payoff from peace must exceed that of an attack. In this case, it must be true that a weak party strictly prefers not to attack.

When a party is expected to attack with positive probability, conflict causes the observers to improve their assessment of the likely attacker's ability (in (3) $\sigma_t^i(1) \geq \sigma_t^i(0)$). For the same reason, peace causes the observers to lower their assessment of the party that was expected to attack. This provides a further incentive against deception, if one were needed. By secretly provoking an attack, a party hands its rival an immediate improvement in its reputation.

The fact that conflict improves the interim reputation of the attacker may suggest that, relative to unknown abilities, we might expect to see more attacks to prove people wrong. This is not the case, however:

Proposition 5 (Less Conflict to Prove People Wrong) *The set of prior reputations that support conflict to prove people wrong in the first period is smaller when parties know their own ability.*

If a party pools on attacking, whether conflict occurs reveals nothing about the ability of either party. Conflict is always expected by the observers. In (3), $\sigma_1^i(0) = \sigma_1^i(1) = 1$, and so $\mu_1^{iC} = \mu_0^i$. Conflict destroys the first period surplus, by the outcome still yields information, causing the observers to update parties reputations according to (6) and (7).

If peace prevails instead, observers would view it as the result of an agency problem. For any $\varepsilon^i > 0$, the probability that party i fails to attack in this equilibrium is $\varepsilon^i(1-\eta^i)$. Observers use (2) to develop interim beliefs, yielding:

$$\mu_1^{iP} = \frac{\mu_0^i \varepsilon^i (1-\eta^i)}{\mu_0^i \varepsilon^i (1-\eta^i) + (1-\mu_0^i)\varepsilon^i(1-\eta^i)} = \mu_0^i.$$

Taking the limit as $\varepsilon^i \to 0$, observers learn nothing about ability if peace prevails.

For party i to pool on attacking, it must be the case that neither ability prefers to deviate by not attacking, thereby tricking the observers into believing that it suffered agency problems. The equilibrium requires that:

$$\pi(\mu_0^i, \mu_0^j)\,(1+\delta) \leq \delta \left[\left(\tfrac{1}{2} - q\mu_0^j\right) \pi(\mu_1^{iV}, \mu_1^{jD}) + \left(\tfrac{1}{2} + q\mu_0^j\right) \pi(\mu_1^{iD}, \mu_1^{jV}) \right]. \tag{10}$$

Equation (10) has the same interpretation as equation (9), except that a weak party's probability of victory, (5), is substituted in. The left-hand side gives the expected payoff from deviating to not attacking, conditional on being pivotal. The observers view the deviation as an agency problem, and do not alter their beliefs about party i. i receives a share $\pi(\mu_0^i, \mu_0^j)$ of both first and second period surpluses. The right-hand side gives the payoff to a weak party from attacking. For the equilibrium to be supported, this has to yield the weak party more.

Compared to unknown abilities, the range of reputations under which they always choose to attack declines. Weak parties know that they are less likely to be victorious than an uninformed party. Their expected payoff from attacking is lower. In order to induce an attack, they must receive an even smaller share of the surplus from peace. However, as no information is revealed when the party attacks, the payoff to peace is unaffected. So for any prior reputation of party j, there are fewer reputations that will induce a pooling attack by party i.

A similar logic holds for peace to prevail. A tough i is more optimistic about their likelihood of winning a fight compared to when they are uninformed. As such, they are more inclined to deviate from an equilibrium in which nobody attacks. The set of prior reputations that support peace in both periods is smaller.

Proposition 6 (Less Peace) *The set of prior reputations that support peace with probability one in the first period is smaller when parties know their own ability.*

In this equilibrium, the observers only expect peace. Since $\sigma_1^i(0) = \sigma_1^i(1) = 0$, (2) suggests that $\mu_1^{iP} = \mu_0^i$. When peace is observed, no information is revealed. If, on the other hand conflict occurs, the observers write it off as one party or other suffering agency problems. The outbreak of conflict reveals no information about either party: $\mu_1^{iC} = \mu_0^i$.

For the equilibrium to be supported, a tough party i must prefer not to attack. This holds if and only if:

$$\pi(\mu_0^i, \mu_0^j)(1+\delta) \geq \delta \left\{ \left[\tfrac{1}{2} + q\left(1-\mu_0^j\right)\right] \pi(\mu_1^{iV}, \mu_1^{jD}) + \left[\tfrac{1}{2} - q\left(1-\mu_0^j\right)\right] \pi(\mu_1^{iD}, \mu_1^{jV}) \right\}. \tag{11}$$

Again, comparing (11) to (9), the difference comes purely through the probability of winning. Since a tough party i knows that they are tough, they are more optimistic about victory. The left-hand side gives i's equilibrium payoff from not attacking, conditional on being pivotal. The right-hand side gives the expected payoff from deviating. By attacking, i tricks the observers into believing that they suffered an agency problem. Nevertheless victory or defeat are informative. For peace to be sustained, tricking the observers in this way cannot be profitable.

For given prior reputations, the shares each party receives from peace, victory and defeat are identical to those under unknown abilities. However, a tough party i expects victory with a higher probability. They consequently need to receive a larger payoff in order to maintain the peace than an uninformed party. For any prior reputation of party j, there are once again fewer reputations that will induce party i to pool on not attacking.

Figure 2 illustrates the pure strategy equilibria under Nash bargaining with private abilities. For comparison, the boundary with unknown abilities is also included. Once again, the party with the weaker reputation is the attacker. Its rival would find the destruction of period one surplus too costly to induce an attack. The regions in which either party attack is smaller, reflecting the less optimistic outlook of a weak ability party. In order to induce it to attack, it must receive an even smaller share of the surplus from peace than an uninformed party. The region in which peace prevails is also smaller, reflecting the more optimistic outlook of a tough party.

In the darker shaded region of Figure 2, the party with the weaker reputation will neither pool on attack or pool on not attacking. In any pooling equilibrium, only the outcome of conflict reveals information. If, for example, conflict occurs when peace was

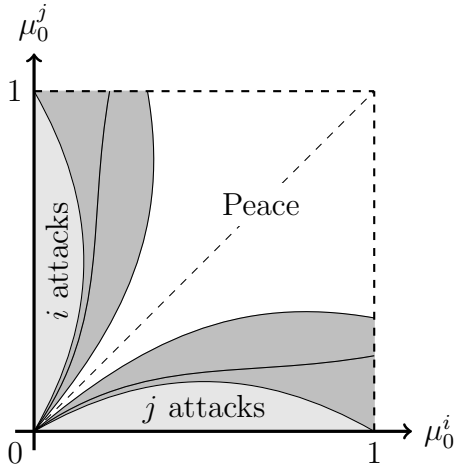

Figure 2: Peace and Proving People Wrong under Nash Bargaining with Private Abilities

expected, they put it down to agency problems. In the shaded region, peace is not possible. A tough party would rather attack, have it written off as an agency problem, and then have the outcome of the conflict announced. Similarly, outright conflict is not possible. A weak party would rather not attack, have it written off as an agency problem, and enjoy the surplus in both periods.

The remaining equilibria all involve some separation between tough and weak parties. As such, whether conflict occurs becomes informative. As the economic intuition behind these equilibria are similar, I will discuss them together, rather than detailing an exhaustive list. At least one such equilibrium exist for any prior reputations, and they take the following form:

Proposition 7 (Conflict is Always Possible to Save Face) *For any prior reputations, there exists an equilibrium in which conflict occurs with positive probability in the first period when parties know their own ability. In such an equilibrium:*

(i) the party with the tougher prior reputation may be the one to attack; and

(ii) observers believe that conflict is more likely as the attacker's prior reputation increases.

Proof. See F. ■

Since these equilibria involve some degree of separation between tough and weak attackers, every outcome occurs with positive probability. Observers are never forced to rely on agency problems to guide their inferences. In line with Propositions 2 and 4, one party will attack with a strictly higher probability when they are tough than when

they are weak. Its rival will not attack. In such equilibria, conflict causes the observers to immediately upgrade the reputation of the attacking party (in (3) $\sigma_1^i(1) > \sigma_1^i(0)$ so $\mu_1^{iC} > \mu_0^i$). This provides an additional incentive to attack. If, instead, peace is observed, the reputation of the party that might have attacked immediately declines. The party is more likely to be weak (in (2) $1 - \sigma_1^i(1) < 1 - \sigma_1^i(0)$ so $\mu_1^{iP} < \mu_0^i$).

This second effect - the decline of reputation when a party does not attack - has the potential to reverse the intuition underpinning the previous results. Until now, it has always been the party with the worse prior reputation that has chosen to attack. It have the least to lose, as the lion's share of the surplus in the current period would go its rival. This is no longer the case. In these equilibria, not attacking is interpreted by the observers as a sign of weakness. Maintaining the status quo may result in the observers giving the majority of the surplus to the rival anyway. Hence even the party who appears to be tougher prior may find itself better off attacking. Conflict is driven by a need to save face.

An increase in the prior reputation of the attacker increases the observer's perceived probability of conflict for two reasons. Firstly, from Proposition 4, tough parties attack with higher probability than weak parties. If the abilities of the attacking party completely separate, then the observers expect to witness an attack in period one with probability μ_0^i. In a semi-separating equilibrium, one ability of the attacker is indifferent between revealing their ability and pooling with the other ability. As the prior reputation of the attacker increases, this indifference is maintained by the ability that mixes attacking with a higher probability. So not only do the observers think it more likely that the attacking party is tough as μ_0^i increases, and is hence more likely to attack, but they expect the tough party to attack more frequently as well.

To expand on the intuition behind this idea, Figure 3 displays the remaining equilibrium under Nash bargaining with private abilities. Given the form of the sharing rule, a weak party would never wish to reveal that it is weak. Doing so would concede the entire surplus to its rival. If such a revelation were possible in equilibrium, a weak party would prefer to mimic a tough party, and maintain the illusion that it might be tough. Consequently, only one semi-separating equilibrium exists. One party mixes between attacking and not if it is tough, and does not attack if it is weak (in the case illustrated, as ever, this is party i). The other party does not attack.

If conflict breaks out, the observers immediately learn that party i is tough. Whilst the outcome of the conflict will have further repercussions for party j's reputation, party i's is fixed: $\mu_1^{iV} = \mu_1^{iD} = 1$. The payoff from conflict for party i is therefore independent

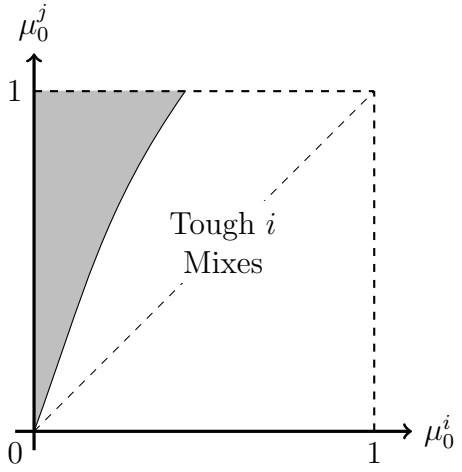

Figure 3: Saving Face under Nash Bargaining with Private Abilities.

of their prior reputation. When attacking, a tough i expects to receive:

$$\delta\left\{\left[\tfrac{1}{2}+q\left(1-\mu_0^j\right)\right]\pi(1,\mu_1^{jD})+\left[\tfrac{1}{2}-q\left(1-\mu_0^j\right)\right]\pi(1,\mu_1^{jV})\right\},$$

conditional on being pivotal.

In order to be indifferent between attacking and defence, a tough i must expect the same payoff from not attacking. We require that:

$$\pi(\mu_1^{iP},\mu_0^j)(1+\delta)=\delta\left\{\left[\tfrac{1}{2}+q\left(1-\mu_0^j\right)\right]\pi(1,\mu_1^{jD})+\left[\tfrac{1}{2}-q\left(1-\mu_0^j\right)\right]\pi(1,\mu_1^{jV})\right\}.$$

Since there are gains from reputation, the payoff from maintaining peace in both periods is increasing in μ_1^{iP}. For any prior reputation of party j, there consequently exists a unique $\mu_1^{iP}(\mu_0^j)$ that maintains a tough party i's indifference. In any such equilibrium, peace must lead the observers to conclude that party i is tough with probability $\mu_1^{iP}(\mu_0^j)$.

If no conflict occurs, observers believe that party i is tough with probability:

$$\mu_1^{iP}=\frac{\mu_0^i\left[1-\sigma_1^i(1)\right]}{\mu_0^i\left[1-\sigma_1^i(1)\right]+\left(1-\mu_0^i\right)}.$$

By mixing with the appropriate probability, the observers can be led develop any interim reputation for i between zero and μ_0^i when peace prevails. In equilibrium, a tough party i attacks with exactly the right probability to generate the $\mu_1^{iP}(\mu_0^j)$ that makes them indifferent.

When is such an equilibrium supported? The payoff party i earns from attacking is less than δ. Since this constitutes less than half the available surplus from peace, it must be that $\mu_1^{iP}(\mu_0^j) < \mu_0^j$. If $\mu_0^i \geq \mu_0^j$, an appropriate probability of attack can always be found to generate the correct posterior reputation from peace. So the equilibrium always exists when party i begins with the tougher prior reputation. In this case, if conflict were not expected, party i would receive the larger share of the peacetime surplus. They would never choose to attack. Conflict is driven solely by the fact that observers downgrade party i's reputation if peace prevails. An attack is expected, and so a tough party i attacks in order to save face. Since a symmetric equilibrium exists in which a tough party j mixes, conflict is always possible when $\mu_0^i < \mu_0^j$ as well.

The boundary of the shaded region in Figure 3 gives the required $\mu_1^{iP}(\mu_0^j)$. On the boundary, a tough party i does not attack. Consequently, $\mu_1^{iP} = \mu_0^i$. They are nevertheless indifferent between revealing themselves to be tough by attacking and maintaining their prior reputation by attacking. Inside the shaded region, party i's prior reputation is too low. A tough party i strictly prefers to attack and reveal themselves to be tough for any feasible reputation from peace. The equilibrium is no longer supported.

Note also that the required μ_1^{iP} is independent of i's prior reputation. If μ_0^i increases, for a given probability of attack, μ_1^{iP} would also increase. A tough i would prefer to deviate and not attack. In order to prevent this, the equilibrium probability that a tough party i attacks increases. This means that, if conflict is not observed, it is more likely that party i is weak. The observers downgrade i's reputation to a greater extent, maintaining the indifference. As party i's prior reputation increases, observers expect more conflict both because it is more likely that party i is tough, and because a tough party i attacks more often.

5 Conclusions

Both throughout history and in contemporary society, conflict has been used to signal violent ability. By developing a fearsome reputation, a party (be it a prince, an army, a general or even a street gang) can ensure the loyalty of the wider population, expanding its sphere of influence at the expense of its rivals. With greater influence comes greater rents.

Such environments generate two motives for conflict, depending upon how much information is conveyed. Fighting creates a victor and a loser. When only this outcome is informative, attacks occur in order to prove people wrong. Parties who have a poor reputation hope that victory will improve their standing with the population, entitling

them to greater rents in the future. Since they earn relatively little by maintaining the status quo, their opportunity cost of preserving the peace is relatively low. In contrast, parties who are already feared have little to gain through further conflict. They already enjoy large rents, and prefer to maintain the peace rather than engage in a costly war that might result in defeat.

When the break out of conflict also signals ability, a party may attack in order to save face. A party that knows that it is tougher than the wider population believes it to be is more optimistic about its chances of victory than a party that knows that it is weaker than the wider population believes. As such, it is more prone to attack. If the population expects this, failing to attack is perceived as a sign of weakness. Loss of face causes the party's sphere of influence to shrink, conceding rents to its rival. Caught between a costly war and a costly loss of face, even parties with ex ante relatively formidable reputations can be induced to attack. Once again, their rival hopes for peace. Peace causes the population previously loyal to their opponent lose faith, and swear fealty to them instead. In this case, conflict is purely driven by the beliefs of the population.

Whilst the results presented seem intuitive, the model used to generate them is relatively simple. By only allowing for two period, the model precludes sophisticated dynamic behaviour. Whilst the destruction of surplus when conflict occurs may be extreme, neither motive for conflict identified depends upon the level cost of conflict. Finally, the assumption that the observers cannot tell directly who started the conflict may rule out a pure signalling motive. If a party knows it is going to be attacked, it may choose to attack as well, in order to improve its interim reputation.

Several extensions suggest themselves. As already noted, the short time horizon prevents credible negotiation between the parties. As the social contract literature shows, the threat of conflict can be used to generate peace given enough time (for example, the Danegeld paid to the Vikings by the Anglo-Saxons in order to prevent attacks). Longer time horizons would also raise possibility of parties investing to improve their ability. For example, countries could elect different leaders (c.f. Jackson and Morelli 2007) or develop new military technology. Moreover, future ability may be related to current reputations, as receiving a higher surplus today provides more resources to invest in improving ability tomorrow. Whilst the current (short-run) reputational motives for conflict would still exist, these additional features would likely give rise to others.

Reputational concerns could also be applied to situations in which rival parties do not engage in conflict with each other. For example, young Roman aristocrats would often spend time at the borders of the empire gaining military victories in order to progress their careers. Whilst they did occasionally fight amongst themselves, they generally

signalled ability by defeating (or not) the Carthaginians, Picts, Huns, Visigoths etc. The need for victories was made more acute by the victories of rival aristocrats in different parts of the Empire.

Acknowledgements

I would like to thank Alwyn Young for his valuable insights and extensive comments. Equally useful have been comments by the editor, the anonymous referees, Chris Colvin, Giacomo De Luca, Iñigo Iturbe-Ormaetxe, Vito Polito and Ákos Valentinyi, as well as seminar participants at Cardiff University and the Long-Run Perspectives on Crime and Conflict conference at Queen's University Belfast.

A Proof of Proposition 1

Proof. Define the discounted payoff to party i in period t be Π_t^i. Suppose that, in equilibrium, party j attacks with probability σ_t^j in period one. For any $\varepsilon^i > 0$ and any $\eta^i > 0$, if party i chooses to attack, their expected payoff is:

$$\begin{aligned}\left(1-\sigma_t^j\right)\left\{\left[1-\varepsilon^i(1-\eta^i)\right]\mathbb{E}_t\left[\Pi_t^i|a_t^i=1,a_t^j=0\right]+\varepsilon^i(1-\eta^i)\mathbb{E}_t\left[\Pi_t^i|a_t^i=a_t^j=0\right]\right\}\\ +\sigma_t^j\mathbb{E}_t\left[\Pi_t^i|a_t^j=1\right]\end{aligned}$$

If they chooses not to attack, their expected payoff is:

$$\begin{aligned}\left(1-\sigma_t^j\right)\left\{\varepsilon^i\eta^i\mathbb{E}_t\left[\Pi_t^i|a_t^i=1,a_t^j=0\right]+\left(1-\varepsilon^i\eta^i\right)\mathbb{E}_t\left[\Pi_t^i|a_t^i=a_t^j=0\right]\right\}\\ +\sigma_t^j\mathbb{E}_t\left[\Pi_t^i|a_t^j=1\right]\end{aligned}$$

So party i will choose to attack if and only if:

$$\begin{aligned}
&\left(1-\sigma_t^j\right)\left\{\left[1-\varepsilon^i(1-\eta^i)\right]\mathbb{E}_t\left[\Pi_t^i|a_t^i=1,a_t^j=0\right]+\varepsilon^i(1-\eta^i)\mathbb{E}_t\left[\Pi_t^i|a_t^i=a_t^j=0\right]\right\}\\
&\qquad+\sigma_t^j\mathbb{E}_t\left[\Pi_t^i|a_t^j=1\right]\\
&\quad\geq\left(1-\sigma_t^j\right)\left\{\varepsilon^i\eta^i\mathbb{E}_t\left[\Pi_t^i|a_t^i=1,a_t^j=0\right]+\left(1-\varepsilon^i\eta^i\right)\mathbb{E}_t\left[\Pi_t^i|a_t^i=a_t^j=0\right]\right\}\\
&\qquad+\sigma_t^j\mathbb{E}_t\left[\Pi_t^i|a_t^j=1\right]\\
\iff&\left[1-\varepsilon^i(1-\eta^i)\right]\mathbb{E}_t\left[\Pi_t^i|a_t^i=1,a_t^j=0\right]+\varepsilon^i(1-\eta^i)\mathbb{E}_t\left[\Pi_t^i|a_t^i=a_t^j=0\right]\\
&\qquad\geq\varepsilon^i\eta^i\mathbb{E}_t\left[\Pi_t^i|a_t^i=1,a_t^j=0\right]+\left(1-\varepsilon^i\eta^i\right)\mathbb{E}_t\left[\Pi_t^i|a_t^i=a_t^j=0\right]\\
\iff&\left(1-\varepsilon^i\right)\mathbb{E}_t\left[\Pi_t^i|a_t^i=1,a_t^j=0\right]\geq\left(1-\varepsilon^i\right)\mathbb{E}_t\left[\Pi_t^i|a_t^i=a_t^j=0\right]\\
\iff&\mathbb{E}_t\left[\Pi_t^i|a_t^i=1,a_t^j=0\right]\geq\mathbb{E}_t\left[\Pi_t^i|a_t^i=a_t^j=0\right].
\end{aligned}$$

Party i will only attack in each period if and only if it is a best response, conditional on party j not attacking, and party i not suffering an agency problem. Each party chooses their action under the assumption that they are pivotal. This completes the proof. ■

B Proof of Corollary 1

Proof. From Proposition 1, each party decides whether to attack by conditioning on their rival not attacking and their not suffering an agency problem. In this case, if party i attacks, the surplus is destroyed, and neither party receives a payoff. If he does not attack, he receives $\pi(\mu_2^{iP},\mu_2^{jP})>0$. So party i will not attack. A symmetric argument holds for party j, and so a fight only occurs in the second period as the result of an agency problem. This completes the proof. ■

C Proof of Proposition 2

Proof. If party j chooses to attack in period one, then they destroys the period one surplus. For this to be (weakly) profitable, it must be the case that:

$$\pi(\mu_1^{jP},\mu_1^{iP})(1+\delta)<\delta.$$

The payoff that party j receives should no fight occur in either period must be less than the (discounted) surplus that remains if they were to attack in period one. Since the

sharing rule is efficient, it must be that:

$$\begin{aligned}&\left[1-\pi(\mu_1^{iP},\mu_1^{jP})\right](1+\delta)<\delta\\&\iff \pi(\mu_1^{iP},\mu_1^{jP})(1+\delta)>1>\delta.\end{aligned}$$

Party i gains a higher payoff by not attacking than they would from attacking, even if that resulted in him securing the entire period two surplus. It cannot be profitable for them to attack. This completes the proof. ■

D Proof of Corollary 2

Proof. Suppose that $\mu_1^{iP} > \mu_1^{jP}$. If party i chooses to attack, it must be that:

$$\pi(\mu_1^{iP},\mu_1^{jP})(1+\delta)<\delta.$$

In this case, by Proposition 2, party j would never choose to attack since:

$$\pi(\mu_1^{jP},\mu_1^{iP})(1+\delta)>\delta.$$

Combining these two inequalities, it must be the case that:

$$\pi(\mu_1^{iP},\mu_1^{jP})<\pi(\mu_1^{jP},\mu_1^{iP}).$$

However, since there are gains to reputation, we have that:

$$\pi(\mu_1^{iP},\mu_1^{jP})>\pi(\mu_1^{jP},\mu_1^{jP})=\tfrac{1}{2}.$$

So, by efficiency, $\pi(\mu_1^{jP},\mu_1^{iP}) < \frac{1}{2}$. Party i receives a higher share of the surplus if no fight occurs. This yields a contradiction, and completes the proof. ■

E Proof of Proposition 4

Proof. For a weak party i, the expected payoff from attacking assuming no agency problem is:

$$\delta\left[\left(\tfrac{1}{2}-q\mu_1^{jP}\right)\pi(\mu_1^{iV},\mu_1^{jD})+\left(\tfrac{1}{2}+q\mu_1^{jP}\right)\pi(\mu_1^{iD},\mu_1^{jV})\right],$$

whereas for a tough party i it is:

$$\delta\left\{\left[\tfrac{1}{2}+q\left(1-\mu_1^{jP}\right)\right]\pi(\mu_1^{iV},\mu_1^{jD})+\left[\tfrac{1}{2}-q\left(1-\mu_1^{jP}\right)\right]\pi(\mu_1^{iD},\mu_1^{jV})\right\}.$$

Subtracting the former from the latter yields:

$$\delta q\left[\pi(\mu_1^{iV},\mu_1^{jD})-\pi(\mu_1^{iD},\mu_1^{jV})\right]>0.$$

A tough party i always expects a higher payoff from attacking than a weak party. Conversely, if they do not attack, conditional on being pivotal, both abilities receive $\pi(\mu_1^{iP},\mu_1^{jP})$. The expected gain from attacking is higher for a tough party than a weak party. So if a weak party attacks with positive probability, a tough party strictly prefers to attack. This completes the proof. ■

F Proof of Proposition 7

Proof. Consider the following strategy profile. Party i does not attack when they are weak and attacks with probability $\sigma_1^i(1)$ when they are tough. Party j never attacks. If conflict occurs, observers conclude that party i attacked, and hence must be tough: $\mu_1^{iC}=1$. Party j's reputation is unaffected. If peace prevails, then the observers conclude that either party i is weak, or that they are tough and randomly did not attack. i's interim reputation becomes:

$$\mu_1^{iP}=\frac{\mu_0^i\left[1-\sigma_1^i(1)\right]}{\mu_0^i\left[1-\sigma_1^i(1)\right]+(1-\mu_0^i)}. \tag{12}$$

i's reputation suffers when they do not attack.

For a tough party i to mix between attacking and not, they must be indifferent between the two actions. So:

$$\pi(\mu_1^{iP},\mu_0^j)\,(1+\delta)=\delta\left\{\left[\tfrac{1}{2}+q\left(1-\mu_0^j\right)\right]\pi(1,\mu_1^{jD})+\left[\tfrac{1}{2}-q\left(1-\mu_0^j\right)\right]\pi(1,\mu_1^{jV})\right\}.$$

Note, however, that the right-hand side of the above equation is independent of μ_0^i. Since attacking reveals i to be tough, their reputation is unaffected by victory or defeat. So, for each μ_0^j, the above equation defines a unique $\mu_1^{iP}(\mu_0^j)$ that maintains a tough party i's indifference. $\mu_1^{iP}(\mu_0^j)<\mu_0^j$, since it must be the case that $\pi(\mu_1^{iP}(\mu_0^j),\mu_0^j)\,(1+\delta)\leq\delta$ and so $\pi(\mu_1^{iP}(\mu_0^j),\mu_0^j)<\frac{1}{2}$.

In order to complete the proof, I show that for any $\mu_0^i>\mu_1^{iP}(\mu_0^j)$, we can find a $\sigma_1^i(1)\in$

$(0,1)$ that generates an interim reputation of $\mu_1^{iP}(\mu_0^j)$ following peace. Rearranging (12) to solve for $\sigma_1^i(1)$ yields:

$$\sigma_1^i(1) = \frac{\mu_0^i - \mu_1^{iP}(\mu_0^j)}{\mu_0^i\left(1 - \mu_1^{iP}(\mu_0^j)\right)}.$$

So long as $\mu_1^{iP}(\mu_0^j) < \mu_0^i < 1$ the equilibrium is supported. In particular, it is supported for any $\mu_0^i \geq \mu_0^j$. Moreover, since $\mu_1^{iP}(\mu_0^j) < \mu_0^j$, it must be that a tough party i attacks with strictly positive probability over this range. The probability of attack is also increasing in μ_0^i. So conflict is always possible when $\mu_0^i \geq \mu_0^j$.

Since an identical equilibrium exists in which a tough party j mixes and party i does not attack, conflict is also possible whenever $\mu_0^i \leq \mu_0^j$. For any prior reputations, an equilibrium exists in which conflict occurs with strictly positive probability. This completes the proof. ■

References

References

Atakan, A.E., Ekmekci, M., 2012. Reputation in long-run relationships. The Review of Economic Studies 79, 451–480.

Azam, J.P., Mesnard, A., 2003. Civil war and the social contract. Public Choice 115, 455–475.

Baliga, S., Sjöström, T., 2004. Arms races and negotiations. The Review of Economic Studies 71, 351–369.

Baliga, S., Sjöström, T., 2008. Strategic ambiguity and arms proliferation. Journal of Political Economy 116, 1023–1057.

Bar-Isaac, H., 2003. Reputation and survival: Learning in a dynamic signalling model. The Review of Economic Studies 70, 231–251.

Bester, H., Wärneryd, K., 2006. Conflict and the social contract. Scandinavian Journal of Economics 108, 231–249.

Brito, D.L., Intriligator, M.D., 1985. Conflict, war, and redistribution. The American Political Science Review 79, 943–957.

Bueno de Mesquita, B., Morrow, J.D., Zorick, E.R., 1997. Capabilities, perception, and escalation. The American Political Science Review 91, 15–27.

Chassang, S., Padró i Miquel, G., 2010. Conflict and deterrence under strategic risk. The Quarterly Journal of Economics 125, 1821–1858.

Crost, B., Felter, J., Johnston, P., 2014. Aid under fire: Development projects and civil conflict. American Economic Review 104, 1833–1856.

Edmond, C., 2013. Information manipulation, coordination, and regime change. The Review of Economic Studies 80, 1422–1458.

Fearon, J.D., 1995. Rationalist explanations for war. International Organization 49, 379–414.

Fudenberg, D., Levine, D.K., 1989. Reputation and equilibrium selection in games with a patient player. Econometrica 57, 759–778.

Fudenberg, D., Levine, D.K., 1992. Maintaining a reputation when strategies are imperfectly observed. The Review of Economic Studies 59, 561–579.

Gambetta, D., 1996. The Sicilian Mafia: The Business of Private Protection. Harvard University Press, Cambridge, MA.

Garfinkel, M.R., 1990. Arming as a strategic investment in a cooperative equilibrium. The American Economic Review 80, 50–68.

Garfinkel, M.R., Skaperdas, S., 2000. Conflict without misperceptions or incomplete information: How the future matters. The Journal of Conflict Resolution 44, 793–807.

Garfinkel, M.R., Skaperdas, S., 2007. Economics of conflict: An overview, in: Sandler, T., Hartley, K. (Eds.), Handbook of Defense Economics. North-Holland, Amsterdam. volume 2. chapter 22, pp. 649–709.

Glaeser, E.L., 2009. The political economy of warfare, in: Hess, G.D. (Ed.), Guns and Butter: The Economic Causes and Consequences of Conflict. MIT Press, Cambridge, MA. CESifo Seminar Series, pp. pp. 33–74.

Grossman, H.I., 1991. A general equilibrium model of insurrections. The American Economic Review 81, 912–921.

Grossman, H.I., Kim, M., 1995. Swords or plowshares? a theory of the security of claims to property. Journal of Political Economy 103, 1275–1288.

Hirshleifer, J., 1989. Conflict and rent-seeking success functions: Ratio vs. difference models of relative success. Public Choice 63, 101–112.

Hirshleifer, J., 1991. The technology of conflict as an economic activity. The American Economic Review 81, 130–134.

Jackson, M.O., Morelli, M., 2007. Political bias and war. The American Economic Review 97, 1353–1373.

Jankowski, M.S., 1991. Islands in the Street: Gangs and American Urban Society. University of California Press, Berkeley and Los Angeles, CA.

Jia, R., Liang, P., 2014. A game of thrones: Power structure and the stability of regimes, in: Wärneryd, K. (Ed.), The Economics of Conflict: Theory and Empirical Evidence. MIT Press, Cambridge, MA. CESifo Seminar Series, pp. 79–104.

Konrad, K.A., Skaperdas, S., 1997. Credible threats in extortion. Journal of Economic Behavior & Organization 33, 23–39.

Konrad, K.A., Skaperdas, S., 1998. Extortion. Economica 65, 461–477.

Kreps, D.M., Wilson, R., 1982. Reputation and imperfect information. Journal of Economic Theory 27, 253 – 279.

Levitt, S.D., Venkatesh, S.A., 2000. An economic analysis of a drug-selling gang's finances. The Quarterly Journal of Economics 115, 755–789.

Machiavelli, N., 1532. The prince, in: Skinner, Q., Price, R. (Eds.), Machiavelli: The Prince. Cambridge University Press, Cambridge, UK. Cambridge Texts in the History of Political Thought.

Milgrom, P., Roberts, J., 1982. Predation, reputation, and entry deterrence. Journal of Economic Theory 27, 280–312.

Selten, R., 1978. The chain store paradox. Theory and Decision 9, 127–159.

Shapiro, C., 1982. Consumer information, product quality, and seller reputation. The Bell Journal of Economics 13, 20–35.

Silverman, D., 2004. Street crime and street culture. International Economic Review 45, 761–786.

Skaperdas, S., 1996. Contest success functions. Economic Theory 7, 283 – 290.

Sun Tzu, c. 500 B.C.E. The art of war, in: Conners, S., Giles, L. (Eds.), The Art of War by Sun Tzu - Classic Edition. El Paso Norte Press.

Tadelis, S., 1999. What's in a name? reputation as a tradeable asset. The American Economic Review 89, 548–563.

Tadelis, S., 2002. The market for reputations as an incentive mechanism. Journal of Political Economy 110, 854–882.

Tullock, G., 1967. The welfare costs of tariffs, monopolies, and theft. Western Economic Journal 5, 224–232.

Tullock, G., 1980. Efficient rent-seeking, in: Buchanan, J.M., Tollison, R.D., Tullock, G. (Eds.), Towards a Theory of the Rent-Seeking Society. Texas A & M University Press, College Station, TX, pp. 97–112.

Wärneryd, K., 2003. Information in conflicts. Journal of Economic Theory 110, 121–136.